**The role of Learning Analytics in Educational Data Mining**

Shalini G1, Robinson Joel M2,B.Muthazhagan3

1Research Scholar, PRIST University, 2,3Kings Engineering College, Chennai

[1nirashakthi@gmail.com](mailto:1nirashakthi@gmail.com) 2 [joel@kingsedu.ac.in](mailto:joel@kingsedu.ac.in), 3muthazhaganb@gmail.com

**Abstract:-EDM (Educational Data Mining) is a data mining algorithm that is used in the field of education. Data mining is a multidisciplinary field that combines artificial intelligence, machine learning, statistics, and database systems to produce results. Data mining may be used in a number of ways in the educational industry to improve student performance and educational institution reputation. Data gathering can indeed be applied to a broad variety of applications in the education system for the aim of boosting student performance and also the status of academic institutions, thanks to the emergence of new technologies that can harness data children's educational and the massive expansion in data sources.EDM refers to the procedures for analysing data from the learning environment in order to have a better understanding of students in their learning environment. The ability for teachers to gain real-time insight about the performance of students, including those who are at risk, can be extremely beneficial in arranging educational activities. It can be motivating and encouraging for students to get information about their performance in comparison to their peers or their progress toward their specific goals. Learning environment data has grown in size in recent years, necessitating the employment of Big Data technologies and methods to manage it. Databases that do not employ typical SQL-based queries are used to address this problem. Data is compressed at rest and in memory using compression algorithms. Before being aggregated, data is broken into smaller pieces and analysed by a huge number of devices spread around the network.**

Keywords: EDU, SQL, Big data, artificial intelligence

**I. Introduction**

Despite the vast variety and volume of data available, many organisations are attempting to derive value from it [1]. The fields of Learning AnalyticsLA and Education Data Mining , in particular, are envisioned as a collection of behaviours that produce knowledge artefacts through reified and codified actions. These communities engage with one another and share information [2]. Netflix, for example, analyses viewer and viewing patterns, as well as the likes and dislikes of popular shows, and uses this information to generate recommendations. In the future, we may be able to leverage big data from DNA to find the best treatment for genetic disorders like cancer. The usage of digital tools and technology in education results in vast amounts of data being collected from students.

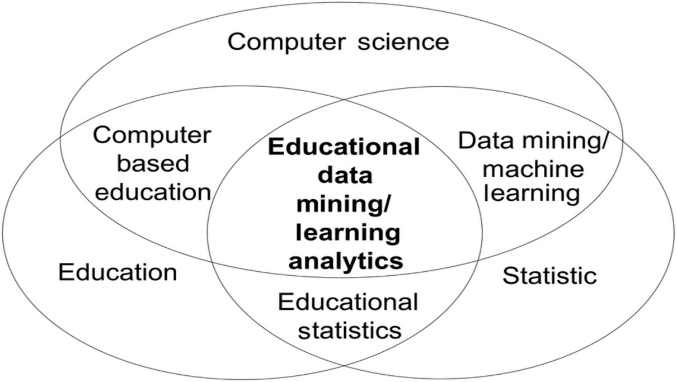


Fig 1.1 Relationship of EDM and LA

In transportation, there are various tactics and challenges, as well as the use of technology such as PCs, smart rooms, projectors, and eBooks. However, these resources are useful only when we know which understudy need which type of resource or, in other words, if we can predict the outcomes of understudies, we can improve results and decrease dropout rates [10].

The field of analytics that uses data collected by students to anticipate educational outcomes in order to adapt instruction. 2015, Junco and Clem In 2015, Xing et al. stated the same thing. “Learning analytics is the measurement [4], collection, analysis and reporting of data learners and their contexts, for purpose of understanding and optimising learning and the environments in which it occurs”-From LAK 2011, conference webpage

The availability of real-time knowledge into the performance of learners, including at-risk kids, can be a great aid in preparing teaching activities for instructors. Pupils might be motivated and encouraged by receiving information and feedback in relation to their classmates or their development toward specific goals. In today's world of higher education, administrators and decision-makers face enormous uncertainty as a result of budget cuts and worldwide competitiveness. Learning analytics is the application of analytics to learning data. The types of data analytics [3] can be applied to learning analytics. What happened when descriptive data analytics were used? The second is diagnostic - what went wrong? The third is prognostic – what will occur? Finally, there's the directive – how do we make it happen?

Researchers frequently choose the following variables to study students' performance: The first is the student's previous grades. Standardised test results, such as college admission scores, are examples of scores. Internal examinations, model exams, and university level exams are only some of the tests available to pupils. They compute the high school GPA and the Graduate Record Examination score based on the exam. The second factor is school characteristics such as the kind of school and the school district. And the administration's official rating of the institution. The third kind is demographic data on students, such as their age, gender (male or female), ethics, and so on. Finally, there is the academic performance of the student, which includes test results. Homework scores and final exam gpa scores are used to assess a student's performance at home.Scores on trail courses and technical tests can also be helpful indicators. Other indications could be derived from questionnaires [8].

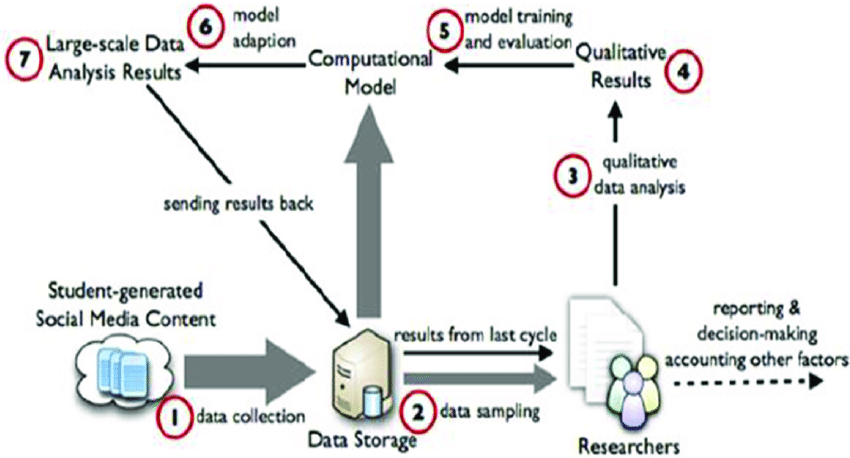


Fig 1.2 Architecture of the Process of Learning Analysis

Descriptive analytics is defined as the process of describing or summarising raw data in order to make it clear, i.e. it gives a summary of facts and figures in a comprehensible format. It will also aid in the discovery of trends in previous data. It also responds to the inquiry, "What went wrong?" It's a sort of graph used in education to demonstrate how well pupils performed on various examinations. It's also utilised to see how a student's performance changes between two tests. In the process of utilising descriptive statistics to analyse a training set of educational data gathered from many sources in blended learning settings. The calculated variety in values [7] demonstrated the utility of the histogram discretization approach during the pre-processing phase.

Diagnostic Analytics is deals with the process of finding out the reason that have led to an event in the past by making use of descriptive analytics. It provides a solution to the inquiry, "Why did things happen the way they did?". Predictive Analytics is the prediction which is done based on the data from past events [5]. It provides a response to the inquiry, "What's next?". Machine Learning are used like Naïve Bayes, SVM, Decision tree etc...

Prescriptive analytics is a method of analysing data and making immediate recommendations for improving the learning process. For example, an instructor may utilise prescriptive analytics to find out that the majority of students needed to take a prerequisite course before enrolling in a newly launched advanced course.

Analyzing and predicting the academic performance of students in higher educational systems and institutions is thought to be extremely important and necessary for the development of a nation in general, as well as for students, instructors, parents, governments, education policymakers, higher education administrative bodies, and higher educational institutions in particular [9].

There are numerous Open Source Tools accessible. There are a number of open source technologies [5] that can assist in taming Big Data. The following are some of the most useful tools. MongoDB is a platform-independent document-oriented database management system. Instead of a table-based design, it uses JSON-like documents. Hadoop is a distributed processing system that distributes large datasets over clusters of networked computers using basic programming techniques. Hadoop employs the Map Reduce programming style and architecture. It allows enormous clusters of computing nodes to simultaneously handle massive amounts of data. Orange is one of the Python-based programme language that used for analysing and mining large amounts of data. It has a simple user interface, drag-and-drop functionality, and a number of add-ons. Weka is a large-scale data processing system based on Java. It has a large number of algorithms that may be used to extract information from data.

**Table:1Literature learning**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Year | Author | Technique | Performance | Remark |
| 2021 | Samira ElAtia[1] | From a large scale, big data educational research viewpoint, aim to deliver relevant and substantial answers on how both LA and EDM may benefit higher education. | Analysing numerous projections about the impact of EDM on future educational settings, particularly in light of the current scenario, which has switched education worldwide to some type of eLearning paradigm. | Raise concerns about basic ethical and privacy considerations when employing EDM and LA in higher education. |
| 2021 | David J.Lemay,ClareBaek,TenzinDoleck[2] | An exploratory retrospective design using research abstracts as the canonical item created across all disciplinary boundaries to seek to understand disciplinary distinctions between the LA and EDM groups. | The use of advanced statistical learning techniques to extract meaningful insights from massive data streams for better teaching and learning appears to be gaining traction in the field of educational research. | Uneven access tends to benefit a select few at the expense of the majority. |
| 2019 | Hanan Aldowah, Hosam Al-Samarrie,Wan Mohamed Fauzy[3] | There were four key learning dimensions identified: predictive, behavioural, and visualisation.. | Data mining techniques are being used across these dimensions, as shown in the diagram. | EDM and LA in higher education can help build a student-centered approach and provide critical tools for institutions to employ in their quest of continuous improvement. |
| 2014 | Papamitsiou.Z&Economides.A[4] | They examine and interpret the findings of the gathered research using non-statistical approaches. | Its goal is to give the reader a thorough overview of current information on Learning Analytics (LA) and Educational Data Mining, as well as its influence on adaptive learning. | During the last six years (2008-2013), the literature on experimental case studies undertaken in the domain was reviewed. |
| 2015 | Katrina Sin and Loganathan Muthu [5] | The classroom was built around three different models: behavioural, cognitive, and constructivist. | It takes a look at some of the most current Big Data applications in education and provides an overview of the literature on Educational Data Mining and Learning Analytics. | The traditional learning environments have gradually mutated into community based learning environment. |

**II. Proposed Work**

So far, we've looked at theoretical comparisons, several forms of analytics, literature reviews from previous years, and an overview of big data collection technology. The execution of practical records into machine learning algorithms utilising learning analytics will be the subject of a future paper. The data will be extracted, and any of the algorithms will be used to produce the desired outcome.

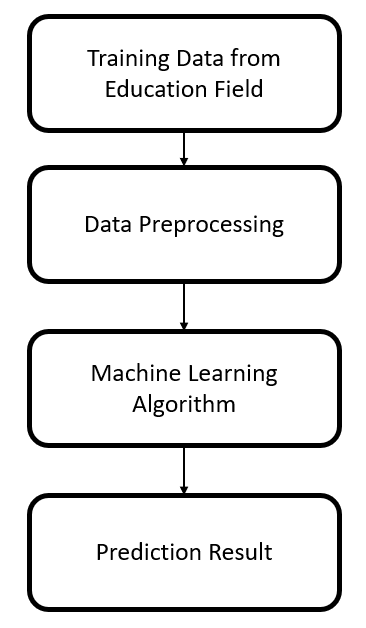


Figure 3. Architecture of proposed system

1. **Conclusion**

The role of Learning Analytics in Educational Data Mining is briefly explained in the study. The information is utilised to evaluate student achievement, teacher performance, and organisational development. It assists teachers in analysing pupils based on data received from the students' module of choice. This information is also utilised to assess the pupils' knowledge. Teachers use the data from the analysis to excel in the most popular topic and identify students' areas of interest. As a result, Learning Analytics is the best method for analysing data and revising the outcome.

**IV. References**

[1] Donald Ipperciel(York University, Canada) and Samira ElAtia (The University of Alberta, Canada). Advancing the Power of Learning Analytics and Big Data in Education.

[2] David J. Lemay, Clare Baek, Tenzin Doleck. Comparison of learning analytics and educational data mining: A topic modelling approach, Computers and Education: Artificial Intelligence, Volume 2. 2021, 100016, ISSN 2666-920X,

[3] Hanan Aldowah, Hosam Al-Samarraie, Wan Mohamad Fauzy, Educational data mining and learning analytics for 21st century higher education: A review and synthesis, Telematics and Informatics, Volume 37, 2019, Pages 13-49, ISSN 0736-5853,

[4] ZacharoulaPapamitsiou, and Anastasios A. Economides. “Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence.” *Journal of Educational Technology & Society*17, no. 4 (2014): 49–64.

[5] Sin, K., Malaysia, O. U., Muthu, L., & University, B. (2015). "APPLICATION OF BIG DATA IN EDUCATION DATA MINING AND LEARNING ANALYTICS – A LITERATURE REVIEW ". ICTACT Journal on Soft Computing, 05(04), 1035–1049.

[6] C. C. Ukwuoma, C. Bo, I. A. Chikwendu and E. Bondzie-Selby, "Performance Analysis of Students Based on Data Mining Techniques: A Literature Review," 2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON), Bangkok, Thailand, 2019, pp. 1-5, doi: 10.1109/TIMESiCON47539.2019.9024396.

[7] C. Vieira, A. J. Magana and M. Boutin, "Using Computational Methods to Analyze Educational Data," 2019 IEEE Frontiers in Education Conference (FIE), Covington, KY, USA, 2019, pp. 1-4, doi: 10.1109/FIE43999.2019.9028599.

[8] Y. Yao, Z. Zhang, H. Cui, T. Ren and J. Xiao, "The Influence of Student Abilities and High School on Student Growth: A Case Study of Chinese National College Entrance Exam," in IEEE Access, vol. 7, pp. 148254-148264, 2019, doi: 10.1109/ACCESS.2019.2946503.

[9] A. O. Enaro and S. Chakraborty, "A Review on the Academic performance analysis of higher Education Students using Data Mining Techniques," 2019 International Conference on contemporary Computing and Informatics (IC3I), Singapore, 2019, pp. 256-260, doi: 10.1109/IC3I46837.2019.9055632.

[10] M. Ayyavaraiah, "Analysis of a Data Mining Based Student Data Collection," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 1-6, doi: 10.1109/ICCMC.2019.8819716.